Testes de Hipóteses

Curso de Introdução à Econometria usando o R

Vítor Wilher

analisemacro.com.br

1 de Dezembro de 2016

Plano de Voo

O que vimos até aqui

1. Apresentação do Curso
2. Apresentação do R
3. Introdução ao mundo da econometria
4. Mínimos Quadrados Ordinários (MQO) como uma ferramenta algébrica
5. O modelo de regressão linear
6. Qualidade do ajuste do modelo linear

O que veremos hoje

1. Introdução
2. Um teste t simples
3. Testando uma restrição linear
4. O Teste F
5. Testando múltiplas restrições lineares com o R
6. Tipos de erro e p-valor

Introdução

Na seção 05 vimos que era preciso fazer uma suposição explícita sobre a distribuição dos termos de erro, de modo a podermos construir uma inferência estatística a partir de uma amostra com N observações.[1] Verificamos, então, que

$$\varepsilon \sim N(0, \sigma^2 I_N), \tag{1}$$

significando que o vetor de termos de erro ε possui uma distribuição normal com vetor médio zero e matriz de covariância $\sigma^2 I_N$.

[1] Seção baseado em Verbeek (2012)

Introdução

Uma forma alternativa de representar essa equação pode ser dada por

$$\varepsilon \sim NID(0, \sigma^2), \quad (2)$$

o que é um modo curto de dizer que os termos de erro ε_i são extraídos de forma independente de uma distribuição normal (n.i.d.) com média zero e variância σ^2. Apesar dos termos serem não observáveis, isso não significa que estejamos livres para fazer qualquer suposição que queiremos. Por exemplo, se assumimos que os termos de erro seguem uma distribuição normal, isso implica que y_i (dado um valor para x_i) também segue uma distribuição normal. Claramente, podemos imaginar uma série de variáveis que não segue uma distribuição normal, o que implica que a suposição de normalidade para os termos de erro é inapropriada. Felizmente, nem todas as suposições são igualmente cruciais para a validade dos resultados que seguem e, ademais, a maioria das suposições pode ser empiricamente testada.

Introdução

Para tornar as coisas simples, vamos considerar a matriz X fixa e determinística ou, alternativamente, vamos trabalhar condicionamente sobre os resultados de X. Assim os resultados que seguem se sustentam. Sob as premissas de Gauss-Markov complementadas pela premissa de normalidade (33), o estimador de MQO b é normalmente distribuído com vetor médio β e matriz de covariância $\sigma^2(X'X)^{-1}$, isto é,

$$b \sim N(\beta, \sigma^2(X'X)^{-1}) \tag{3}$$

A prova disso segue diretamente do resultado que b é uma aproximação linear de todos os termos de erro ε_i.

Introdução

O resultado acima implica que cada elemento em b é normalmente distribuído, por exemplo

$$b \sim N(\beta_k, \sigma^2 c_{kk}), \tag{4}$$

onde, como anteriormente, c_{kk} é o elemento (k, k) em $(X'X)^{-1}$. Nós podemos, a propósito, utilizar esse resultado para desenvolver testes de hipóteses sobre o parâmetro populacional β desconhecido. De (45) segue que a variável

$$z = \frac{b_k - \beta_k}{\sigma \sqrt{c_{kk}}} \tag{5}$$

possui uma distribuição normal padrão (i.e., uma distribuição normal com média zero e variância igual a 1). Se substituirmos σ por sua estimativa s, isso não é mais exatamente verdadeiro.

Introdução

Pode ser demonstrado que o estimador não viesado s^2 definido em (31) é independente de b e possui uma distribuição Chi-quadrado com $N - K$ graus de liberdade. Em particular,

$$\frac{(N-K)s2}{\sigma^2} \sim \chi^2_{N-K} \tag{6}$$

Consequentemente, a variável aleatória

$$t_k = \frac{b_k - \beta_k}{s\sqrt{c_{kk}}} \tag{7}$$

será a razão entre a variável normal padrão e a raiz quadrada de uma variável Chi-quadrado independente e, portanto, segue segue uma distribuição t de Student com $N - K$ graus de liberdade.

Introdução

A distribuição t é próxima à distribuição normal padrão, exceto que ela possui caudas mais *gordas*, particularmente quando o número $N-K$ é pequeno. Quanto maior for $N-K$, mais próxima a distribuição t será de uma distribuição normal padrão.

Um teste t simples

O resultado acima pode ser utilizado para construir estatísticas de teste e intervalos de confiança. A ideia geral de um **teste de hipóteses** é a que segue. Partindo de uma **hipótese nula**, é calculada uma estatística de teste que possui uma distribuição conhecida sob a suposição de que a hipótese nula é válida. Em seguida, é decidido se o valor calculado da estatística de teste não provém dessa distribuição, o que indica que a hipótese nula não se sustenta. Para ilustrar, suponha que que nós temos uma hipótese nula que que especifica o valor de β_k, isto é, $H_0 : \beta_k = \beta_k^0$, onde β_k^0 é um valor específico escolhido pelo pesquisador.

Um teste t simples

Se a hipótese é verdadeira, sabemos que a estatística

$$t_k = \frac{b_k - \beta_k^0}{se(b_k)} \tag{8}$$

tem uma distribuição t com $N - K$ graus de liberdade. Se a hipótese nula é falsa, a hipótese alternativa $H_1 : \beta_k \neq \beta_k^0$ se mantém. A quantidade na equação acima, por suposto, é uma estatística de teste que é calculada a partir da estimativa b_k, seu erro padrão $se(b_k)$ e o valor da hipótese β_k^0 sob a hipótese nula.

Um teste t simples

Se o teste estatístico realiza um valor que é muito diferente sob a distribuição nula, nós rejeitamos a hipótese nula. Nesse caso isso significa um valor absoluto para t_k muito grande. Para ser preciso, rejeita-se a hipótese nula se a probabilidade de observarmos um valor de $|t_k|$ muito grande é menor do que um dado **nível de significância**, em geral 5%. A partir disso, podemos definir **valores críticos** $t_{N-K;\alpha/2}$ usando

$$P(|t_k| > t_{N-K;\alpha/2}) = \alpha \tag{9}$$

Um teste t simples

Para $N - K$ não tão pequeno, esses valores críticos são apenas levemente maiores do que aqueles provenientes de uma distribuição normal padrão, para o qual as duas caudas dos valores críticos para $\alpha = 0.05$ será 1.96. Consequentemente, ao nível de 5% a hipótese nula será rejeitada se $|t_k| > 1.96$.

Um teste t simples

O teste acima é nomeado como um **teste de dois lados** porque a hipótese alternativa permite valores de β_k em ambos os lados de β_k^0. Ocasionalmente, a hipótese alternativa possui apenas um lado, por exemplo: o salário esperado para um homem é maior do que aquele esperado para uma mulher. Formalmente, nós definimos a hipótese nula como $H_0 : \beta_k \leq \beta_k^0$ com a alternativa $H_1 : \beta_k > \beta_k^0$.

Um teste t simples

Depois, nós consideramos a distribuição do teste estatístico t_k na fronteira da hipótese nula (isto é, $\beta_k = \beta_k^0$, como antes) e rejeitamos a hipótese nula se t_k for muito grande.[2] Números negativos grandes para t_k são compatíveis com a hipótese nula e não levam a sua rejeição. Assim, para um teste de um lado o valor crítico é determinado como

$$P(t_k > t_{N-K;\alpha/2}) = \alpha \tag{10}$$

[2] Note que valores grandes para b_k levam a valores grandes para t_k.

Um teste t simples

Usando a aproximação normal padrão de novo, nós rejeitamos a hipótese nula ao nível de 5% se $t_k > 1.64$. Pacotes estatísticos tipicamente reportam, por suposto, o seguinte valor de t:

$$t_k = \frac{b_k}{se(b_k)} \tag{11}$$

às vezes referido como *razão t*, que é o ponto estimado dividido por seu erro padrão. A razão t é a estatística t utilizada para testar a hipótese nula que $\beta_k = 0$. Se ela for rejeitada, dizemos que b_k é estatisticamente significativo ou signficativamente diferente de zero. Se uma variável explanatória é estatisticamente significativa isso não significa, entretanto, que haja alguma interpretação econômica. Às vezes, particularmente com conjuntos de dados muito grandes, podemos encontrar significância estatística ao mesmo tempo que a magnitude econômica do seu efeito seja muito pequeno.

Um teste t simples

Por fim, um **intervalo de confiança** pode ser definido como o intervalo de todos os valores de β_k^0 para os quais a hipótese nula $\beta_k = \beta_k^0$ não é rejeitada pelos testes t. Falando livremente, dado uma estimativa b_k e seu erro padrão associado, um intervalo de confiança dá uma classe de valores que provavelmente contém os verdadeiros valores de β_k. Isso deriva do fato que a seguinte inequação se mantém com probabilidade $1 - \alpha$:

$$-t_{N-K;\alpha/2} < \frac{b_k - \beta_k}{se(bk)} < t_{N-K;\alpha/2}, \tag{12}$$

Ou

$$b_k - t_{N-K;\alpha/2} se(b_k) < \beta_k < b_k + t_{N-K;\alpha/2} se(b_k), \tag{13}$$

Um teste t simples

Consequentemente, usando a aproximação normal padrão, um intervalo de confiança de 95% ($\alpha = 0.05$) para β_k será dado pelo intervalo

$$[b_k - 1.96se(b_k), b_k + 1.96se(b_k))] . \tag{14}$$

Em amostras repetidas, 95% desses intervalos irá conter o verdadeiro valor de β_k que é fixo porém um número desconhecido.

Exemplo do Wooldridge

Vamos pegar um exemplo do Wooldridge (2013) para ilustrar melhor o teste t. Na videoaula pegamos os dados de forma on line, exatamente da mesma forma que fizemos nas seções anteriores. O arquivo `gpa1.dta` foi baixado e lido com a função `read.dta`, do pacote `foreign`.

Exemplo do Wooldridge

Uma vez que tenhamos o objeto gpa1 no nosso *Environment*, é possível construir uma regressão que associa a *nota média em um curso superior* à *nota média do ensino médio* e à *nota do teste de avaliação de conhecimentos para ingresso em curso superior*. Nosso conjunto de dados refere-se a uma amostra de 141 estudantes de uma grande universidade dos Estados Unidos. Na videoaula, criamos essa regressão no R. Vamos entendê-la?

Testando uma restrição linear

O teste t visto acima envolve a restrição de um único coeficiente. Às vezes, entretanto, pode ser preciso verificar uma restrição que envolva mais de um coeficiente, como por exemplo $\beta_2 + \beta_3 + ... + \beta_k = 1$. De forma geral, podemos formular uma **hipótese linear** como

$$H_0 : r_1\beta_1 + ... + r_k\beta_k - r^{'}\beta = q \tag{15}$$

para alguma valor escalar q e um vetor K-dimensional r. Nós podemos testar a hipótese acima usando o resultado de que $r^{'}b$ é BLUE para $r^{'}\beta$ com variância $V(r^{'}b) = r^{'}V(b)r$. Substituindo σ^2 na matriz de covariância $V(b)$ pela estimativa s^2 produz uma matriz de covariância estimada, titulada $\hat{V}(b)$.

Testando uma restrição linear

Consequentemente, o erro padrão da combinação linear $r^{'}b$ será $se(r^{'}b) = \sqrt{r^{'}\hat{V}(b)r}$. Como b é K-variável normal, $r^{'}b$ também será normal, de modo que teremos

$$\frac{r^{'}b - r^{'}\beta}{se(r^{'}b)} = t_{N-K} \tag{16}$$

o que é uma generalização direta da equação (48) da apostila.

Testando uma restrição linear

O teste estatístico para H_0 segue

$$t = \frac{r^{'}b - q}{se(r^{'}b)} \tag{17}$$

o qual possui uma distribuição t_{N-K} sob a hipótese nula. Ao nível de 5%, o valor absoluto de t que excede 1.96 leva à rejeição da hipótese nula. Essa forma representa, assim, uma versão mais geral do teste t.

Computar o erro padrão $se(r^{'}b)$ requer a matriz de covariância estimada do vetor b. Infelizmente, alguns pacotes estatísticos não proveem uma forma fácil de obter o erro padrão da estastística t de forma direta.

Testando uma restrição linear

Nesses casos, um modo conveniente de obter o mesmo teste estatístico é via reparametrização do modelo original, tal que a restrição linear H_0 corresponda a uma restrição usual, tal qual $\beta_k^* = 0$. Por exemplo, considere

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \tag{18}$$

e suponha uma restrição como $\beta_2 = \beta_3$. Assim, podemos reescrever o modelo como

$$y_i = \beta_1 + (\beta_2 - \beta_3) x_{i2} + \beta_3 (x_{i3} + x_{i2}) + \varepsilon_i \tag{19}$$

ou

$$y_i = \beta_1 + \beta_2^* x_{i2} + \beta_3 (x_{i3} + x_{i2}) + \varepsilon_i \tag{20}$$

Testando uma restrição linear

Da definição do MQO como minimização do somatório ao quadrado dos resíduos, segue que isso não varia para reparametrização linear. Consequentemente, o estimador de OLS para β_3 em ambas as formulações do modelo será idêntico e o estimador para β_2^* será dado por $b_2 - b_3$. A vantagem da reparametrização é que a hipótese nula pode ser escrita como uma restrição zero no coeficiente da regressão, isto é, $H_0 : \beta_2^* = 0$. Consequentemente, ela pode ser testada utilizando o teste t padrão para β_2^* no modelo reparametrizado, dado por

$$t = \frac{b_2^*}{se(b_2^*)} = \frac{b_2 - b_3}{se(b_2 - b_3)} \tag{21}$$

onde b_2^* é o estimador OLS para β_2^*.[3]

[3] Veja um exemplo no R aqui.

O Teste F

Um teste padrão disponível na maior parte dos pacotes estatísticos é o **teste F**, que irá verificar a hipótese conjunta de que todos os coeficientes, exceto o intercepto, são iguais a zero, isto é,

$$H_0 : \beta_{K-J+1} = ... = \beta_k = 0 \tag{22}$$

A hipótese alternativa nesse caso é que H_0 não é verdadeiro, isto é, ao menos um dos coeficientes não é igual a zero.

A forma mais fácil de computar esse teste é comparar a soma dos resíduos ao quadrado do modelo total com o modelo restrito, isto é, com os J regressores omitidos, tal qual

$$\frac{S_0 - S_1}{\sigma^2} = \chi^2_J \tag{23}$$

Onde S_0 refere-se aos resíduos do modelo restrito e S_1 aos resíduos do modelo total.

O Teste F

Com algumas manipulações, também é possível computar a estatística F utilizando o R^2 da forma abaixo

$$F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)} \tag{24}$$

Onde N é o total de observações e K é o número de regressores.[4]

[4] Para maiores detalhes, ver Verbeek (2012), pg. 27 e 28.

Testando múltiplas restrições lineares com o R

Utilizando a estatística F é possível testar restrições lineares no R de forma bastante simples com a função `linearHypothesis` do pacote `cars`. No nosso exemplo do Wooldridge (2013) visto acima, vamos testar a restrição dada por $hsGPA + ACT = 1$ com o código da videoaula.

Testando múltiplas restrições lineares com o R

Observe que a estatística F é bem maior do que o valor crítico, o que nos leva a rejeitar a hipótese imposta. Ademais, o leitor pode verificar que a função `linearHypothesis` permite combinar múltiplas restrições aplicadas ao mesmo tempo.

Tipos de erro e p-valor

Quando uma hipótese é testada estatiscamente, podemos estar cometendo dois tipos de erro. O primeiro erro é rejeitar a hipótese nula quando a mesma for verdadeira. Chamamos esse erro de **erro do tipo I**. O segundo, chamado de **erro do tipo II** ocorre quando não rejeitamos a hipótese nula quando a alternativa é verdadeira. A probabilidade de um erro de tipo I é controlada pelo pesquisador através de sua escolha do nível de significância α. Quando um teste é realizado ao nível de 5%, a probabilidade de rejeitar a hipótese nula enquanto ela é verdadeira é de 5%. Esta probabilidade (nível de significância) é muitas vezes referida como o tamanho do teste.

Tipos de erro e p-valor

A probabilidade de um erro de tipo II depende dos valores dos parâmetros verdadeiros. Intuitivamente, se o valor verdadeiro desvia-se muito da hipótese nula declarada, a probabilidade de tal erro será relativamente pequena, enquanto que será bastante grande se a hipótese nula estiver próxima da verdade. A probabilidade inversa, ou seja, a probabilidade de se rejeitar a hipótese nula quando ela é falsa, é conhecido como o poder do teste.

Tipos de erro e p-valor

Uma probabilidade final que tem grande impacto em testes estatísticos é usualmente referido como **p-valor**. Significa o mínimo valor para o qual a hipótese nula ainda seria rejeitada. É definida como a probabilidade, sob a hipótese nula, de encontrar uma estatística de teste (em valores absolutos) que exceda o valor da estatística que foi calculada a partir da amostra. Assim, se o p-valor é menor do que o nível de significância α, rejeita-se a hipótese nula. No exemplo do Wooldridge (2013) que colocamos acima, o p-valor do coeficiente ACT foi de 0.38, isto é, há uma probabilidade de 38% de encontrarmos uma estatística de teste maior do que aquela que encontramos, de 0.87, logo não podemos rejeitar a hipótese nula de que o coeficiente é igual a zero, não sendo assim estatisticamente signficativo.

Referências

Verbeek, M. *A Guide to Modern Econometrics*. Editora Wiley, 2012.

Wooldridge, J. M. *Introductory Econometrics: A Modern Approach*. Editora Cengage, 2013.